



ἡ θεός ἀνθρώπων

THEORIA

ISSN 0495-4548 – eISSN 2171-679X

## Justification, conversation, and folk psychology<sup>&</sup>

*(Justificación, conversión y psicología folk)*

VÍCTOR FERNÁNDEZ CASTRO\*

PSL Research University, France

**ABSTRACT:** The aim of this paper is to offer a version of the so-called conversational hypothesis of the ontogenetic connection between language and mindreading (Harris 1996, 2005; Van Cleave and Gauker 2010; Hughes et al. 2006). After arguing against a particular way of understanding the hypothesis (the communicative view), I will start from the justificatory view in philosophy of social cognition (Andrews 2012; Hutto 2004; Zawidzki 2013) to make the case for the idea that the primary function of belief and desire attributions is to justify and normalize deviant patterns of behaviour. Following this framework, I elaborate upon the idea that development of folk psychological skills requires the subjects to engage in conversationally mediated joint and cooperative activities in order to acquire the conceptual capacity of ascribing propositional attitudes. After presenting the general version of the hypothesis, I present several testable sub-hypotheses and some psychological studies that give empirical plausibility to the hypothesis.

**KEYWORDS:** Mindreading, Regulation, Conversation, Language, Social Cognition.

**RESUMEN:** El objetivo de este artículo es ofrecer una versión de la hipótesis conversacional sobre conexión ontogenética entre lenguaje y lectura de mentes (Harris 1996, 2005; Van Cleave and Gauker 2010; Hughes et al. 2006). Después de argumentar contra una versión particular de dicha hipótesis (Versión Comunicativa), partiré de la visión justificativa de la cognición social (Andrews 2012; Hutto 2004; Zawidzki 2013) para defender que la función primaria de la adscripción de creencias y deseos es la de justificar y normalizar patrones de comportamiento anormales. Tomando en consideración este marco teórico, argumentaré que el desarrollo de las habilidades de lectura de mentes requiere que los sujetos participen en actividades cooperativas y conjuntas mediadas conversacionalmente gracias a las cuales pueden adquirir la capacidad conceptual de adscribir actitudes proposicionales. Después de presentar la versión general de la hipótesis, se presentan diferentes sub-hipótesis susceptibles de ser testadas y estudios empíricos que las refuerzan.

**PALABRAS CLAVE:** Lectura de Mentes, Regulación, Conversación, Lenguaje, Cognición Social.

<sup>&</sup> I would like to thank Fernando Martínez-Manrique for their valuable comments and suggestions. This paper was written as part of the projects “Inner speech, Metacognition, and the Narrative View of Identity” (FFI2015-65953-P) and “Contemporary Expressivism and the Indispensability of Normative Vocabulary” (FFI2016-80088-P) of the Spanish Ministry of Economy and Competitiveness.

\* **Correspondence to:** Víctor Fernández Castro. Institut Jean Nicod. Département d'Etudes Cognitives. ENS. PSL Research University. 29 rue d'Ulm 75005 Paris. France – [vfernandezcastro@gmail.com](mailto:vfernandezcastro@gmail.com) – <http://orcid.org/0000-0001-7627-5738>

**How to cite:** Fernández Castro, Víctor. (2019). «Justification, conversation, and folk psychology». *Theoria. An International Journal for Theory, History and Foundations of Science*, 34(1), 73-88. (<https://doi.org/10.1387/theoria.18022>).

Received: 29 August, 2017; Final version: 7 January, 2018.

ISSN 0495-4548 - eISSN 2171-679X / © 2019 UPV/EHU



This article is distributed under the terms of the  
Creative Commons Attribution 4.0 International License

## 1. Introduction

An important part of the research concerning the ontogeny of mindreading has concentrated on how language facilitates the capacity for ascribing mental states (Astington and Baird 2005; De Villiers 2007; Turnbull et al. 2009).<sup>1</sup> These views contrast with theories that claim that mindreading development is independent of language faculty either because such socio-cognitive capacity is partially innate (Carruthers 2013; Fodor 1982) or because its development is independent of language (Perner et al. 2005; Wellman 2014). In this paper, I will focus on a particular view of the connection between language and mindreading, that is, the so-called *Conversational Hypothesis*, according to which children's capacity to assign beliefs and desires to others is tied to the exposure to and engagement with conversational contexts (Harris 1996, 2005; Van Cleave and Gauker 2010).<sup>2</sup> In other words, children acquire the ability to understand others' actions in mental terms when they are exposed to different conversational situations. This theory competes with other approaches to the connection between language and folk psychology; for instance, approaches that highlight the representational capacities of language as the key aspect influencing mindreading abilities (Bermudez 2003; Segal 1998). However, instead of discussing the virtues or problems of the conversational hypothesis in comparison to other theories, this paper brings up different philosophical and empirical issues concerning two ways of understanding the conversational hypothesis. In particular, it discusses how different philosophical approaches to the nature of belief and desire ascriptions result in different interpretations of the conversational hypothesis, and thus, different sub-hypotheses concerning how the relationship between conversation and mindreading must be understood.

Likewise, the aim of this paper is to motivate a version of the conversational hypothesis by considering a particular approach to belief and desire ascriptions. According to this approach, which I will call the *justificatory view*, the primary function of belief and desire ascriptions is not to describe inner psychological states for the sake of prediction or explanation. Instead, propositional attitude ascriptions are responses to counter-normative behaviour which serve, paradigmatically, to justify or condemn particular courses of action. The justificatory view contrasts with another approach we can find in the literature, that is, the *communicative view* (section 2). According to this view, the emergence of belief and desire ascriptions is tied to the practice of talking on the others' behalf. In other words, belief and desire ascriptions are vicarious assertions and commands. After arguing against the communicative view (section 2) and favouring the justificatory view (section 3), I will elaborate on a particular version of the conversational view, according to which the appropriate conversational contexts children must enter into in order to

<sup>1</sup> I shall use the terms 'mindreading', 'theory of mind', 'mentalizing' or 'folk psychology' indistinctively to refer to the capacity for making sense of one's and others' behavior in terms of mental states.

<sup>2</sup> Folk psychology includes several capacities including attributing different psychological states; for instance, emotions, perceptions, or intentions. However, many scholars restrict their analysis to the attribution of propositional attitudes. For example, although Nichols and Stich (2003) recognize the role of emotions in our folk psychological capacities, the attribution of propositional attitudes (desires and beliefs) is, in their view, the primary explanatory target. I follow those authors and restrict my investigation to the attribution of desires, beliefs and other propositional attitudes, including some empirical literature concerning the phenomenon.

develop ascriptional skills are conversationally mediated joint activities and coordinated actions where participants monitor, adjust, sanction and regulate each other's actions (section 3). This version can be instantiated with different sub-hypotheses. In section 4, I present some of these sub-hypotheses and review some empirical findings that speak in favour of their plausibility.

## *2. The Communicative Version of the Conversational Hypothesis*

Before introducing the communicative version of the hypothesis, a succinct definition of the conversational hypothesis is necessary. Broadly considered, defenders of the conversational hypothesis claim that being exposed to and engaging in conversational exchanges is a requirement for acquiring the ability to understand actions in terms of mental states (Beliefs and Desires). In its paradigmatic characterization (Harris 1996, 2005), folk psychological understanding requires the recognition of others' perspectives or points of view. The main feature of mentalizing is to appreciate that others have a different perspective on reality. This feature of mentalizing, Harris argues, is acquired by being exposed to different exchanges of information, where children are reminded that speakers have different beliefs, desires, and intentions. For instance, exposure to conversational exchanges helps children to appreciate the different perspective mothers intend to instil in them: "it is the mother's pragmatic intent, notably her efforts to introduce varying points of view into a given conversation, that is the underlying and effective source of variation" (Harris 2005, 77). Harris claims folk psychologists acquire the sensitivity to understand others as recipients of information. Although Harris presents different empirical results to support his claim, he does not specify the way conversational exchanges precondition belief and desire attributions. In other words, we need to specify which conditions the developmental setting of the conversational exchange needs to meet in order to be a relevant situation for the acquisition of belief and desire ascriptions.

In order to introduce these conditions, Van Cleave and Gauker (2010; see also Gauker 2003) begin from a particular perspective of belief and desire ascriptions. According to this perspective, the primary use of ascriptions is to carry out vicarious speech acts, i.e., ascribing the belief that P has to a subject S is asserting P on the behalf of S; ascribing the desire that Q has to a subject S is commanding Q on the behalf of S. Van Cleave and Gauker introduce two scenarios to exemplify the view:

Scenario One: Billy and Sally are playing with their toys in the living room. Mother is expecting guests in half an hour. Sally steps into the kitchen where her mother is preparing food. Mother says to Sally, 'You and Billy, please pick up all your toys in the living room and take them to your own rooms'. Sally returns to the living room and says to Billy, 'Mom wants us to put away the toys'.

Scenario Two: Billy and Sally are playing in the yard. They expect to attend a friend's birthday party later in the afternoon, but they are not sure when it starts. They see another friend, Markie, at the end of the block. Sally walks down the street to ask when the party starts. When she returns, she says to Billy, 'Markie thinks the party starts at four o'clock'. (Van Cleave and Gauker 2010, 310)

These vicarious speech acts are the basic function of ascriptions that, afterward, can be used to explain behaviour. In fact, Van Cleave and Gauker claim that children can provide par-

tial explanations based on these uses before they engage in complete explanation with reference to mental states.

Given that, the communicative version of the conversational hypothesis can be characterized as claiming that children acquire the ability to attribute mental states in conversational contexts where they can recognize the credibility and authority of others. If the primary function of belief ascriptions is to talk on the other's behalf, then belief attribution requires identifying others as credible spokespersons. Similarly, if the primary function of desire ascriptions is to command on the other's behalf, desire ascriptions require identifying authority. Children must understand that others may have more credibility or authority than they have, so they can voice this credibility and authority in the practice of contributing information to a cooperative plan (Van Cleave and Gauker 2010, 318). Likewise, recognizing others as being credible or having authority is the key to producing vicarious speech acts, and therefore, to acquiring the primary capacities for ascribing propositional attitudes. Although Van Cleave and Gauker do not elaborate upon which types of conversational environments are required, they point to disputes as one example of conversation that helps to exercise these primary uses of belief and desire. In fact, they consider the research by Bartsch and Wellman (1995, 112-126) as evidence in support of their position. In these studies, Bartsch and Wellman discovered that early uses of belief and desire seem to be bound to situations of solving disputes, rather than explaining actions. These uses, Van Cleave and Gauker argue, can often be viewed as indirect discourse attributions or paraphrasing attributions.

In a similar vein, Tooming (2016) claims that a virtue of the communicative view is not only that it can explain the reason why the concepts of belief and desire emerge (extending conversational practice); but it also explains how this linguistic practice accounts for cases where people use belief and desire ascriptions to emphasize different points of view or even to damage others' social statuses. For instance, when we treat someone as if they had asserted something ridiculous (Tooming 2016, 198). Although Tooming aligns himself with the position of Van Cleave and Gauker, he maintains that the linguistic function of belief and desire ascriptions serve mostly to emphasize the agreement/disagreement and approval/disapproval between the attributer and the attributee. These functions, he argues, cannot be displayed without the possession of mental concepts, and thus, it gives a plausible explanation of how a community of agents with linguistic capacities would acquire the mental concepts necessary to be able to talk on others' behalf, and emphasizes these attitudes.

Be that as it may, the virtue of the communicative view is that it provides a framework from which we can understand the role of conversations in acquiring mentalizing abilities. If the primary function of ascribing beliefs and desires is to talk on others' behalf, we need to acquire conversational competencies to be able to attribute mental states. Furthermore, the communicative view can account for several empirical findings, including the correlational studies that emphasize the connection between disputes and success in false-belief tasks (Brown et al. 1996; Foote and Holmes-Lonergan 2003; Slomkowski and Dunn 1992; see section 4).<sup>3</sup> However, the communicative view faces three important challenges.

<sup>3</sup> Originally designed by Wimmer and Perner (1983), the false-belief task (FB-task) was designed to tap mindreading abilities. In this task, a child is exposed to a scenario where a character, Maxi, puts choc-

First, although van Cleave and Gauker recognize that children start using mental-state verbs in first-personal terms (with what they call an ‘expressive function’), they consider the uses as derivative, that is, they are not serious reflections of genuine mental reference. However, they do not offer any reason for why we should not take these uses into consideration as cases of mental ascriptions. These first-personal ascriptions serve to modulate or indicate different grades of commitment to a particular assertion<sup>4</sup> (It’s raining, I think); that is, they have a contrastive or pragmatic function (Bartsch & Wellman 1995; Urmson 1952; Wierzbicka 2006). Certainly, Van Cleave and Gauker could argue these uses of mental verbs do not reflect the natural function of mental verbs. Although this interpretation is quite common, it could be biased by the fact that psychologists considered that the primary function of mental predicates is anticipating and explaining actions by referring to psychological states, which is precisely the interpretation that the defenders of the communicative view aim to undermine.

Furthermore, we have reasons to assume that first-personal explanation in terms of mental states remains a pervasive explanatory tool in adulthood. Several empirical findings in social psychology speak in favour of this point. For instance, Malle et al. (2007; see also Malle et al. 2001) carried out intensive research providing empirical evidence demonstrating different asymmetries in how people explain behaviour depending on whether they are actors or observers of the action. In the context of this research, they found that, in general, actors produce many more reason and mental-state explanations than observers in comparison with causal history or trait explanations. In other words, intentional explanations in mental terms are usually provided to explain one’s own behaviour rather than others’. Certainly, Van Cleave and Gauker could argue that we must distinguish between first and third-personal explanation at this point. After all, the purpose of their view is to account for how our capacity to interpret others appears in ontogeny. Certainly, many philosophers consider first and third-personal mental explanations to follow different developmental paths and even that one perspective precludes the other.<sup>5</sup> However, I believe that an alternative explanation that could account for both ranges of cases has the potential to be a better alternative to the communicative view (see section 3). Thus, one must take into account

---

olate into a box x. When Maxi is not present, his mother displaces the chocolate from x into a box y. Children have to indicate the box where Maxi will look for the chocolate when he returns. Only when the child is able to represent Maxi’s wrong belief, is he able to point correctly to box x. This task tests whether children have an «explicit representation of the wrongness of this person’s belief in relation to one’s own knowledge» (Wimmer and Perner 1983, 103). They found that younger children (3-years old) frequently fail in this task. These results have been consistently reproduced. Furthermore, different versions of the task were proposed (Astington and Jenkins 1999; Gopnik and Astington 1988; Perner et al. 1987).

<sup>4</sup> Although these uses are usually presented as having a pragmatic function, rather than a mentalizing function, there is no reason why one may not use this attenuation of commitments as a way to explain the motives of an action (see section 3).

<sup>5</sup> I am aware that there are reasons to believe that either the first-person perspective or the third person perspective appears first in development, and even that one could be a requirement to the other (see Carruthers 2011; Goldman 2006). I do not consider these reasons to be conclusive but evaluating them goes beyond the purpose of this paper. For the sake of the argument, I take for granted that any position that can account for the development of both perspectives at the same time is in a better position than its contenders, at least for reasons of economy of explanation.

these uses of first-personal ascriptions as positive evidence which the different contenders must account for. Summing up, first-personal uses of mental terms appear first in development and they are quite pervasive in intentional explanation, so they seem a good candidate to reflect primary uses of belief and desire ascriptions which the communicative view fails to account for.

A second challenge to the communicative view comes from considerations about the two preconditions the theory imposes on engaging in vicarious speech acts: authority and credibility. In principle, recognizing authority does not seem to be a necessary condition for attributing desires. For instance, children usually attribute desires to peers and toys (My cookie monster wants my mommy to make some cookies for him), who can hardly be considered as sources of authority. In fact, if, as Van Cleave and Gauker argue, the appearance of ascriptions is connected to contexts of dispute, then one may expect children to understand conflicting desires as a precondition to understanding authority and not the other way around. On the other hand, considering credibility as a precondition of belief attribution is also problematic. The challenge has to do with the idea that vicarious speech acts preclude explanations. Although Van Cleave and Gauker indicate that “explanatory attributions of belief often occur in the context of excusing a mistake” (2010, 316), they claim that vicarious assertions appear ontogenetically before this explanatory function. However, it is hard to see why children could not exculpate actions before they can produce vicarious speech acts. After all, mental states can be recruited as mitigators for exculpating mistakes (Kitty ate her cookie because she thought it was her cookie) and, as we have seen before, these pragmatic functions of mental states seem to be the first demonstrations of mental vocabulary.

Finally, the communicative view must address a problem with some developmental findings. An ontogenetic consequence of the communicative view is that vicarious speech acts appear in contexts of dispute, and thus, Van Cleave and Gauker argue, one may expect children to start using mental ascriptions in such contexts. However, this claim seems to be too restrictive. The empirical literature on the connection between conversations and mindreading shows that the type of contexts that facilitate mindreading abilities are broader; including pretend play (Astington and Jenkins 1999; Hughes et al. 2006), cooperative activities (Huyder et al. 2017) or story-telling (Nelson 2006). Thus, this communicative version of the hypothesis falls short of covering the range of conversational contexts that seem to be connected to mental ascriptions.

In summary, the communicative view faces several difficulties. Firstly, the theory disregards the pervasiveness of first-personal uses of mental verbs and their role in the development of mentalizing skills. Secondly, credibility and authority seem to be too demanding as preconditions of mental-state acquisition. And thirdly, the predictions produced by the hypothesis at the developmental level are very restricted, ignoring several conversational contexts that seem to be important in the development of ascriptional capacities. In the following section, I will offer an alternative to the communicative view. This alternative, which I will call the *justificatory view*, starts with the idea that the primary function of belief and desire ascriptions is justifying or advocating a particular action. Furthermore, I shall explore the preconditions that this view imposes on specifying the role that conversations play in the acquisition of mentalizing skills.

### 3. *What is a Belief for? The Justificatory View*

In recent years, several authors have questioned the idea that our anticipatory capacities and sense-making practices rely on the ability to attribute mental states. On the contrary, they argue that social cognition is based on a myriad of socially shared normative structures and routines that facilitate our coordination and understanding (Andrews 2012; Gallagher 2001; Gallagher and Hutto 2008; Hutto 2004; McGeer 2001, 2007, 2015; Zawidzki 2013). According to this proposal, our anticipatory capacities do not require attributing mental states. Instead, “the work is done and carried by the world, embedded in the norms and routines that structure such interactions” (McGeer 2001, 119). We make sense of each other because we share a set of routines, normative structures and patterns of behaviour that make our actions comprehensible to others; so these cultural and rational norms, complex cultural patterns of behaviour and routines, enhance cooperation and facilitate the performance of structurally complex social projects. But they also provide structures that help us to consider what is acceptable or normal in particular situations (Hutto 2004, 560). In other words, our capacity to make sense of each other requires understanding a different array of normative structures that indicate which behaviour is in question in a particular situation.

The question is which role propositional attitude ascriptions play in our social capacities within this framework. From this view, folk psychological ascriptions are bound to the elucidation of counter-normative behaviour (Andrews 2009, 444-446; Fernández Castro 2017, 8-9; Hutto 2004, 560-562; Zawidzki 2013, 216- 222). As Andrews puts it:

Given an understanding of norms in a society, and the ability to recognize and sanction violations, there developed a need to understand actions that violated the norms. Explanations for norm-violating behavior that didn't cite a person's reasons either led to excluding the individual (e.g., “He fed because he is crazy, so let's stop sharing meat with him”), or they failed to satisfy those who demand an explanation. This need to have a satisfactory reason for the behavior of one's companions is what drives the need to develop another sort of explanation, namely reason explanations. There is a significant benefit to being able to explain behavior that violates norms, because explanations of the right sort can also serve to justify behavior. (Andrews 2009, 445)

In contexts where social understanding is governed by norms, it makes sense to have the possibility of justifying behaviour and making it understandable in order to avoid public sanctions. Ascriptions serve to rationalize our behaviour when it is perceived as anomalous or when we are encouraged to exculpate ourselves. Rationalizing our actions in terms of mental states allows us to both avoid and anticipate possible sanctions in contexts where our behavioural patterns are open to sanction by our peers. The use of mental-state ascriptions, is thus restricted to cases where the actions under explanation are perceived by the attributer as anomalous or counter-normative. In other words, evaluating a person as hoping, desiring or believing is to specify a certain responsibility and significance in order to justify, exculpate or condemn their behaviour when their actions violate certain expectations.

A reason to support this view comes from the everyday observation that humans engage in social explanations of this type when they need to assign certain responsibilities to a subject in order to rationalize their behaviour. One way to recognize this point is by considering how difficult people find it to explain actions they strongly condemn (Andrews 2012, 155), in particular when these actions are accounted for in terms of reasons or ascriptions. For instance, when someone offers an explanation for immoral behaviour, e.g.,

'Daesh killed hundreds because of Western countries' foreign policy in their territories'; people tend to interpret the explainer to be a supporter or defender of the immoral action. Conjectures aside, this view coheres with the aforementioned studies by Malle et al. (2007). Firstly, as the studies suggest, one may expect people to provide more reasons in terms of mental states when they are the actors because they likely aim to advocate their own actions. For instance, introducing first-personal mental verbs to attenuate their commitment with the reason motivating the action (*I arrived late because the meeting was at eight, I thought*). Furthermore, one may also expect from the justificatory view that people will introduce third-personal attribution when attempting to put others' actions in a positive light. Other empirical studies supporting the view are those of Korman and Malle (2016; see also Uttich and Lombrozo 2010) which have shown that people offer many more reasons in terms of mental states when they face puzzling actions, in contrast to ordinary actions. In these experiments, a group of participants was presented with sentences describing situations in which some agents' behaviours were described as "puzzling with respect to social perceivers' prior knowledge and expectancies about behaviour in general" (Korman and Malle 2016, 3), in contrast to another group presented with sentences describing ordinary behaviours. Although both groups provided a similar number of explanations, the group presented with puzzling behaviour tended to provide more reason explanations (in contrast to trait or causal explanations) and mental-state explanations than the other. In other words, subjects provide reason-explanations in mental terms to support others' action when they have to justify it.

This approach to social cognition opens two questions for the purposes of this paper. First, which preconditions does the justificatory view impose on acquiring attributional capacities? And second, what consequences do these preconditions have for the conversational hypothesis? The first question has to do with the circumstances that prompt the necessity of justifying actions. The rationale behind the justificatory view is the idea that we use belief and desire ascriptions in order to exculpate or advocate a particular counter-normative action. In this sense, acquiring the appropriate mental concepts necessitates interactions where the actions of the participants are open to monitoring, sanction and regulation. When other agents expect us to behave according to the norms that police our social interactions, violating or transgressing such norms will trigger regulative responses on the behalf of the other participants. These regulative responses include acts such as: blaming, questioning the behaviour, asking for reasons or verbal sanctions. Belief and desire ascriptions are tools for normalizing our anomalous action in order to avoid this type of regulative response; that is, ascriptions are tools for advocating the action that others perceive as anomalous. As a result, the justificatory view claims an agent must be open to regulative responses in order to develop the capacity for ascribing mental states.

The justificatory view is subject to a possible objection at this point. Tooming (2016) argues that the justificatory view fails to give a plausible account of the rationale behind the emergence of the concept of beliefs and desires. In other words, in a community of agents where the concept of beliefs and desires does not exist, normalizing counter-normative behaviours does not necessarily force the emergence of belief and desire ascriptions because the normalizing function can be carried out by other means. While performing vicarious speech acts, Tooming claims, seems to be a function restricted to belief and desire ascriptions, justifying others' counter normative behaviour is a function that can be satisfied by



other means. For instance, justifying the action through other actions (the action of breaking eggs is justified by the action of making an omelette) or appealing to another norm that justifies the behaviour that could seem counter-normative in the first place (one behaviour could be counter-normative because it violates a politeness rule but it could be normalized by a rationality rule, for instance).

Although Tooming is right in considering that we display non-mentalistic justificatory strategies, there are two justificatory functions that are hard to perform without using mental concepts. Firstly, agents can avoid possible future sanctions by introducing first-personal ascriptions that attenuate their behavioural commitments. For instance, consider an example offered by Zawidzki where a hunter asserts that an animal was to the north when in fact it was not. Tooming is right in considering that justifying the assertion or the action of going north does not require mental states ascriptions. However, the hunter could avoid possible sanction by indicating that he is not fully committed to the assertion that there is prey to the north; namely, saying '*I think the animal was to the north*'. This justification in advance would help the members of the community to avoid possible future sanctions if his audience does not find the prey in the north. Secondly, an agent can justify the action of another person by appealing to a reason she does not share (see Tanney 2013, 133-148). In order to see why, consider these two sentences:

- (1) Sara ran away because the building was on fire.
- (2) Sara ran away because *she thought* the building was on fire.

Sentence (1) does not only imply that the target (Sara) was justified in running away, but also that the ascriber is committed to the building being on fire. On the contrary, sentence (2) has the same justificatory function of the action of Sara but it does not commit the speaker to the building being on fire. In this sense, introducing the mental concept specifies the ascriber's position toward reason.<sup>6</sup> Likewise, the justificatory view can point to at least two justificatory functions of mindreading that could not be carried out without acquiring mental concepts and thus, the justificatory view is well posed to elucidate the preconditions for acquiring mental concepts. That is, social interactions where the participants are open to regulation and sanctions that therefore impose the necessity of providing explanation and justifications so they may be avoided.

The second question the justificatory version of the conversational hypothesis must answer is which consequences the aforementioned preconditions impose on the conversational hypothesis. From an ontogenetic point of view, one may hypothesize that children need to engage in conversational contexts where their actions are perceived as anomalous or counter-normative in order to develop the capacity to provide reason-explanation in terms of mental states for the sake of justifying or exculpating these actions. That is, children need to be embedded in social encounters where participants make behavioural de-

---

<sup>6</sup> At this point, one may have noticed that imputing mental ascriptions to others as a way to distance oneself from a reason that otherwise normalizes the action does not substantially differ from talking on the behalf of another person. In fact, these justificatory functions play some of the roles that Tooming and van Cleave and Gauker attribute to mental ascriptions; i.e., excusing mistakes or damaging others' social status. At this point, thus, the difference between the communicative and justificatory view might become blurred. However, the main difference, I believe, relies on the function and the emphasis the two theories assign to first-personal ascription.

viations which are relevant to each other in such a way that the actions trigger reactive or sanctioning responses that the participants want to avoid. In these contexts, agents deploy ascriptions in order to avoid these possible sanctions by exculpating their own or others' behaviour, anticipating possible objections or making explicit evaluations to solve possibly conflicting interpretations. For instance, one may need to appeal to propositional attitudes when someone questions our actions (*Why did you pull the doll from your sister? I thought she wanted to break it*), contexts where someone notices our errors and we need to excuse the mistake (*Why did you leave your brother alone? He did not want to play with me*) or situations where we want to indicate our degree or lack of support concerning an assertion to anticipate possible negative reactions (*The dinosaur, I think, is broken*). In any case, all these examples share the basic function of reacting to possible indications of violations of social norms, that is, the possibility that a certain course of action is contravening a norm. As a result, the consequence of the justificatory view for the conversational hypothesis is that the appropriate conversation exchanges that pave the way to acquiring ascription capabilities are those conversationally mediated joint activities where children are subject to verbal regulative responses, so they can undertake or exempt themselves or the attributee of certain responsibilities.

Given that, we can characterize the justificatory version of the conversational hypothesis as claiming that mindreading abilities are acquired in conversational engagements where the actions of the different participants become relevant to each other in such a way that their actions are open to sanction and regulation. So, mental-state attributions emerge as a tool for justifying or normalizing counter-normative actions. From a developmental point of view, paradigmatic examples of these situations involve cases of reciprocal play, caregiver-child interactions, cooperative games, group task resolutions, or cooperative-pretend play. Certainly, this list is far from being exhaustive, but it provides some clues to the type of developmental environments the conversational hypothesis must focus on. In the following section, in order to provide some plausibility to the justificatory version of the conversational hypothesis, three possible sub-hypotheses are described, along with evidence of their correlation with mindreading success.

#### 4. Pretend Play, Disputes and Moral Transgression

The version of the conversational hypothesis I describe in this paper envisages a developmental connection between conversational exchanges where participants are exposed to regulation and sanction and the appearance of mindreading abilities. This general hypothesis can be subdivided in several sub-hypotheses concerning different situations joining precondition and success in mindreading tasks. In this section, we concentrate on three possible sub-hypotheses and some empirical findings that speak in favour of them. Of course, this evidence is far from being conclusive, but it gives to the general hypothesis an empirical starting point to motivate further research.

The first testable claim the justificatory version of the conversational hypothesis can generate exploits the connection between contexts of disputes and mental state ascriptions. From this view, cases where the agents penalize and question each other's actions are the proper circumstances where mentalizing skills arise. Now, one may expect these responses to generate cases of disagreement and disputes, and thus, cases where agents offer mental-

state explanations to advocate their own point of view. In this sense, the view predicts a correlation between cases of dispute and disagreements and mindreading abilities. Likewise, the justificatory version of the conversational hypothesis shares with Van Cleave's and Gauker's view the emphasis on disputes and arguments as appropriate contexts for acquiring ascriptional skills. The main difference is that while Van Cleave and Gauker consider that disputes facilitate children's capacity to voice others' credibility and authority, the justificatory conversational hypothesis considers that disputes and disagreements correlate with mindreading because disputes arise in circumstances where agents regulate, monitor and sanction each other. Certainly, the justificatory view does not deny that we can voice others' authority or credibility in order to justify a particular action for solving a dispute. However, the emphasis relies on the justificatory function of the ascriptions. In section 2, I mentioned a set of studies showing the correlation between success in false-belief tasks and argumentation and conflicts (Brown et al. 1996; Foote and Holmes-Lonergan 2003; Slomkowski and Dunn 1992). For instance, Foote and Holmes-Lonergan (2003) confirmed that children's use of mental-state terms that occurred during conflicts with siblings was a solid predictor of performance on false-belief tasks. Foote and Holmes-Lonergan (2003) also found that the children who use other-oriented arguments during these conflicts, which involves negotiation or reasoning that incorporates either the partner's interests or the interests of both children, scored higher in false-belief tasks. These findings suggest a strong connection between arguments and justifications and mental-state ascriptions. Similarly, Dunn and colleagues found that children who engage in explanatory discourse in the context of a dispute with their mother and siblings at 36 months were more successful in mental-state understanding at 40 months (Dunn and Brown 1993; Tesla and Dunn 1992; see Dunn 1994, for a review). Admittedly, this evidence is not strong enough to shift the balance on any of the views. However, the justificatory view allows us to look at the connection between ascriptions and other types of situations.

At this point, one may wonder whether or not there is empirical evidence that can decide between the communicative and the justificatory view.<sup>7</sup> In section 2, I have argued that the communicative view does not seem to accommodate the role that first-personal ascriptions with pragmatic function play in the development of mindreading. Furthermore, it is argued that the justificatory view considers that these uses have an exculpatory function for future behaviours. For instance, we use expressions like '*I think*' to mitigate the commitment with the proposition the verb makes (*it's raining, I think*). Following this line, Lewis et al. (2012) have carried out research that aims to show that children's delay in passing the false-belief tasks at four years old is due to their systematic exposure to such pragmatic functions. That is, children are more familiar with the use of sentences where the verb 'think' appears with sentences where the that-clause is true than those where it is false. As they say, children find it difficult to pass standard false-belief tasks because "they tend to underestimate the relevance of belief states, and the probability that a speaker is intending to comment on someone's belief state rather than the state of the actual world" (Lewis et al., 254-255). In their experiments, Lewis et al. (2012) introduces a change-of-location paradigm where the children saw a character (Dora) who is looking for another character (Swipper) in a hide-and-seek game. Dora wrongly thinks Swipper was behind a box when

---

<sup>7</sup> Thanks to an anonymous reviewer at this journal for bringing this to my attention.

she was behind a curtain. A second seeker comes into scene and he rightly thinks that Swiper is behind the curtain. The conflicting beliefs between the seekers help the experimenters to determine whether children tend to interpret the situation as if the verb 'think' were interpreted with a pragmatic function. They found that children underestimate the role of the mental state and they focus on the real state of the world, indicating that the use of the verb 'think' is pragmatic.

A second plausible set of contexts where children may be prompted to display justificatory ascriptions are those involving cooperative interactions. Consider the case of cooperative pretend plays, in which children create different characters (e.g. cowboys, pirates) and take their roles in different pretend situations. An important feature of this type of game is that children negotiate and create different normative structures to determine how the character must behave. So, children embedded in these situations systematically monitor and regulate each other's actions to make sure the characters are properly performed (see Wyman et al. 2009). These situations of monitoring and regulating are the preconditions for the emergence of mental ascriptions according to the justificatory conversational hypothesis. Children regulate each other's actions in a way that prompts the use of justifications and explanations when they monitor each other during play. Now, the connection between pretend play and mindreading has been highly explored in the empirical literature (Nichols and Stich 2003; Leslie 1994), showing a strong correlation between the two. One may object that pretend play is not necessarily performed in cooperation. Children often perform pretend play by themselves. Thus, pretend play cannot be considered supportive evidence for the conversational hypothesis, since they do not always involve social interactions. However, as Harris (2005) has discussed, there are several reasons to assume that it is the social and conversational features of cooperative pretend play that make a difference for acquiring ascriptional abilities. For instance, solitary pretend play does not seem to correlate with social understanding (Harris 2005, 78), while role assignments or the frequency of joint proposal do (Astington and Jenkins 1999; Jenkins and Astington 1996). Furthermore, Hughes and Dunn (1997) showed that that children are more likely to use mental-state terms in the context of pretend and reciprocal play than in other contexts (see also Jenkins and Astington 2000). These findings were replicated by Hughes et al. (2006). During their studies, the experimenters visited 140 families with 2-year-old children and their siblings. The experimenters prompted the subjects to play with costumes and toys that led them to engage in pretend play. After those visits, the children were tested in different mindreading and verbal tasks. Among the results, Hughes et al. (2006) emphasized reciprocal and cooperative pretend play as one of the stronger predictors (along with age and verbal skills) of the use of mental-state terms.

Finally, the justificatory version of the conversational hypothesis can be instantiated with a particular hypothesis concerning the role of mental-state ascriptions in contexts of moral transgression. The justificatory view starts with the idea that the primary function of mental-state ascriptions is to rationalize or exculpate actions that are perceived as counter-normative or anomalous. Given that, one can predict that contexts of transgressions of moral norms may prompt the use of belief and desire ascriptions. Following this line, several studies have shown a strong correlation between the ability to understand other mental states and moral permissibility and sensitivity (Astington 2004; Dunn et al. 2000; Killen et al. 2011). For instance, Dunn and her colleagues found that children with an understanding of mental states (higher scores in theory of mind) were also more

likely to justify the moral transgressions of their friends using arguments incorporating their mental states. Likewise, Killen et al (2014) found that children with a worse understanding of mental states attributed worse intentions to transgressors, indicating they were not able to understand that false-beliefs could provoke non-intentional immoral actions.

Certainly, this evidence is far from conclusive. Even though these studies may speak in favour of the justificatory view of the conversational hypothesis, the general hypothesis and the sub-hypothesis presented in this section require the design of new empirical settings that contribute to reinforce or falsify the hypothesis. In particular, testing the hypothesis would require designing more training studies to show that the correlations presented in this section manifest a causal connection. Furthermore, the correlational studies would demand a control of the linguistic marks of justificatory uses of mental states.

Finally, the justificatory version of the conversational hypothesis also provides an opportunity to reinterpret other findings concerning studies in false-belief tasks. Although children are able to pass explicit versions of these tasks around the age of four, several studies have shown they are able to pass it before when the experimental design does not test predictive capacities, rather than explanatory or justificatory ones (Bartsch and Wellman 1989; Robinson and Mitchell 1995). For instance, Bartsch and Wellman presented 3 year-old children, 4 year-old children and adults with different short descriptive theories they had to explain: "Jane is searching for her kitten under the piano. The kitten is hidden under the chair; why do you think she is searching there?" Then, the different answers were analysed. They found that 3 years old were able to explain anomalous behaviour (behaviour that transgresses normalized patterns of actions) in terms of false belief even when they failed to accomplish predictive versions of the task. These findings reinforce the idea behind the justificatory view, according to which mental concept acquisition is tied to rationalizing and normalizing counter-normative actions, rather than anticipatory capacities. Certainly, this does not mean mental-state attribution cannot serve predictive purposes but it indicates that the primary function of mental states should relate to justificatory or explanatory purposes.

## *5. Conclusions*

New philosophical approaches in social cognition bring to the table different theoretical and methodological issues concerning the variety of strategies and mechanisms underlying our social behaviour. These new approaches provide new opportunities and tools for re-evaluating previous debates concerning social cognition. In this paper, we have concentrated on the consequences of the justificatory view for the developmental connection between language and mental ascriptions.

The justificatory view strongly emphasizes the idea that attributing mental states is not the primary strategy we use to deal with social interactions. Instead, humans exploit different normative strategies to engage with and interpret social situations. The justificatory view starts from the idea that we exploit different normative strategies to make sense of others. In order to maintain those norms and routines, humans display different regulative strategies including giving and asking for reasons, and blaming or sanctioning when someone contravenes norms. This framework promotes the idea that belief and desire as-

criptions serve to rationalize and condemn actions. Attributing desires and beliefs serves to assign certain responsibilities in order to facilitate exculpation (or condemnation) of a particular action.

This framework permits putting forward a proposal concerning the connection between conversation and folk psychology. Given the appropriate contexts where ascriptions are used (those of counter-normative behaviours, anticipating objections and regulating others' actions), one may expect a strong developmental correlation between those conversational contexts and folk psychological skills. According to the justificatory conversational hypothesis, children need to engage in joint activities where regulative responses appear, for instance, cooperative pretend play, cooperative actions, disputes and disagreements, and so on. In other words, interactions where the courses of action of the different parts become salient and relevant to each other in such a way that the participants need to regulate each other according to a range of norms.

This justificatory version of the conversational hypothesis is, then, not only driven by the conceptual and philosophical motivations behind the justificatory perspective, but also by different sources of evidence speaking in its favour. Along with classical studies showing the connection between conversations and folk psychology, there are some studies that speak in favour of this particular version of the conversational hypothesis. In particular, these studies show how pretend play, disputes or moral interactions are the appropriate contexts to acquire the capacity for mental ascriptions.

## REFERENCES

- Andrews, Kristin. 2009. Understanding norms without a theory of mind. *Inquiry* 52/5: 433-448.
- Andrews, Kristin. 2012. *Do apes read minds? Toward a new Folk Psychology*. Cambridge: MIT Press.
- Astington, Janet W. and Jodie A. Baird. eds. 2005. *Why language matters for theory of mind*. Oxford: Oxford University Press.
- Astington, Janet W. and Jennifer M. Jenkins. 1999. A longitudinal study of the relation between language and theory-of-mind development. *Developmental psychology* 35/5: 1311-1320
- Astington, Janet W. 2004. Bridging the gap between theory of mind and moral reasoning. In Jodie A. Baird and Bryan W. Sokol, eds., *Connections between theory of mind and sociomoral development: No. 103. New directions for child and adolescent development*, 63-72. San Francisco, CA: Jossey Bass.
- Bartsch, Karen and Henry Wellman. 1989. Young children's attribution of action to beliefs and desires. *Child Development* 60/4: 946-964.
- Bartsch, Karen and Henry Wellman. 1995. *Children talk about the mind*. Oxford: Oxford University Press.
- Bermudez, José L. 2003. *Thinking without words*. Oxford University Press.
- Brown, Jane N., Nancy Donelan-McCall, and Judy Dunn. 1996. Why talk about mental states? the significance of children's conversations with friends, siblings, and mothers. *Child Development* 67/3: 836-849.
- Carruthers, Peter. 2011. *The opacity of mind: an integrative theory of self-knowledge*. Oxford: Oxford University Press.
- Carruthers, Peter. 2013. Mindreading in infancy. *Mind and Language* 28/2: 141-172.
- De Villiers, Jill G. 2007. The interface of language and theory of mind. *Lingua. International Review of General Linguistics* 117/11: 1858-1878.
- Dunn, Judy. 1994. Changing minds and changing relationships. In Charlie Lewis and Peter Mitchell, eds., *Children's early understanding of mind: Origins and development*, 297-310. London: Psychology Press.
- Dunn, Judy and Jane R. Brown. 1993. Early conversations about causality: Content, pragmatics and developmental change. *British Journal of Developmental Psychology* 11/2: 107-123.

- Dunn, Judy and Penny Munn. 1987. Development of justification in disputes with mother and sibling. *Developmental Psychology* 23/6: 791-798.
- Dunn, Judy, Jane R. Brown, Cheril Slomkowski, Caroline Tesla, and Leslie Young-blade. 1991. Young children's understanding of other people's feelings and beliefs: Individual differences and their antecedents. *Child Development* 62/6: 1352-1366.
- Dunn, Judy, Alexandra L. Cutting, & Helen Demetrious. 2000. Moral sensibility, understanding others, and children's friendship interactions in the preschool period. *British Journal of Developmental Psychology* 18/2: 159-177.
- Fernández Castro, Víctor. 2017, online first. Regulation, Normativity and Folk Psychology. *Topoi. An International Review of Philosophy*. DOI: 10.1007/s11245-017-9511-7.
- Fodor, Jerry 1992. A theory of the child's theory of mind. *Cognition* 44 /3: 283-296.
- Foote, Rachel C. and Heather A. Holmes-Lonergan. 2003. Sibling conflict and theory of mind. *British Journal of Developmental Psychology* 21/1: 45-58.
- Gallagher Shaun 2001. The practice of mind: Theory, simulation or primary interaction? *Journal of Consciousness Studies*, 8/7-7: 83-108.
- Gallagher Shaun and Daniel D. Hutto 2008. Understanding others through primary interaction and narrative practice. In Jordan Zlatev, Tomothy Racine, Chris Sinha, and Esa Itkonen, eds., *The Shared Mind: Perspectives on intersubjectivity*, 17-38. Amsterdam: John Benjamins.
- Gauker, Christopher. 2003. *Words without meaning*. Cambridge: MIT Press.
- Goldman, Alvin I. 2006. *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford: Oxford University Press.
- Gopnik, Alison and Janet W. Astington. 1998. Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, 59/1: 26-37.
- Harris, Peter H. 1996. Desires, beliefs, and language. In Peter Carruthers and Peter K. Smith, eds., *Theories of theories of mind*, 200-220. Cambridge: Cambridge University Press.
- Harris, Peter H. 2005. Conversation, pretense, and theory of mind. In Janet W. Astington and Jodie Baird, eds., *Why language matters for theory of mind*, 70-83. New York: Oxford University Press.
- Harris, Peter H., Marc Rosnay, and Francisco Pons. 2005. Language and children's understanding of mental states. *Current directions in psychological science* 14/2: 69-73.
- Hughes, Claire and Judy Dunn. 1997. "pretend you didn't know": Preschoolers' talk about mental states in pretend play. *Cognitive Development* 12/4: 477-497.
- Hughes, Claire, Keiko K. Fujisawa, Rosie Ensor, Serena Lecce, and Rachel Marfleet. 2006. Cooperation and conversations about the mind: A study of individual differences in 2-year-olds and their siblings. *British Journal of Developmental Psychology* 24/1: 53-72.
- Hutto, Daniel D. 2004. The limits of spectatorial folk psychology. *Mind and language* 19/5: 548-73.
- Huyder, Vanessa, Elisabeth Nilsen and Sarah Bacso. 2017. The relationship between children's executive functioning, theory of mind, and verbal skills with their own and others' behaviour in a cooperative context: Changes in relations from early to middle school-age. *Infant and Child Development* 26/6.
- Jenkins, Jennifer J. and Janet W. Astington. 1996. Cognitive factors and family structure associated with theory of mind development in young children. *Developmental psychology* 32/1: 70-78.
- Jenkins, Jennifer J. and Janet W. Astington. 2000. Theory of mind and social behavior: Causal models tested in a longitudinal study. *Merrill-Palmer Quarterly* 46/2: 203-220.
- Killen, Melanie, Kelly L. Mulvey, Cameron Richardson, Noah Jampol and Amanda Woodward 2011. The Accidental Transgressor: Morally Relevant Theory of Mind. *Cognition* 119/2: 197-215.
- Korman, Joanna and Bertram F. Malle. 2016. Grasping for traits or reasons? how people grapple with puzzling social behaviors. *Personality and Social Psychology Bulletin* 42/11: 1451-1465.
- Leslie, Alan M. 1994. Pretending and believing: Issues in the theory of tomm. *Cognition* 50: 211-238.
- Lewis, Shevaun, Jeffrey Lidz, and Valentine Hacquard. 2012. The semantics and pragmatics of belief reports in preschoolers. *Proceeding Semantics and Linguistic Theory* 22: 247-267.

- Malle, Bertram M., Joshua Knobe, and Sarah Nelson. 2007. Actor-observer asymmetries in explanations of behavior: New answers to an old question. *Journal of Personality and Social Psychology* 93: 491-514.
- McGeer, Victoria 2001. Psycho-practice, psycho-theory and the contrastive case of autism. how practices of mind become second-nature. *Journal of Consciousness Studies* 8/5-6: 109-132.
- McGeer, Victoria 2007. The regulative dimension of folk psychology. In Daniel D. Hutto and Matthew Ratcliffe, eds., *Folk Psychology re-assessed*, 137-156. Kluwer/Springer Press.
- McGeer, Victoria 2015. Mind-making practices: The social infrastructure of self-knowing agency and responsibility. *Philosophical Explorations* 18/2: 259-281.
- Nichols, Shaun and Stephen Stich. 2003. *Mindreading. An integrated account of pretence, self-awareness, and understanding other minds*. Oxford: Oxford University Press.
- Perner, Josef, Susan R. Leekam and Heinz Wimmer. 1987. Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology* 5/2: 125-137.
- Perner, Josef, Petra Zauner and Manuel Sprung. 2005. What Does "That" Have to Do with Point of View? Conflicting Desires and "Want" in German. In Janet W. Astington and Jodie Baird, eds., *Why language matters for theory of mind*, 220-244. New York: Oxford University Press.
- Robinson, Elisabeth J. and Peter Mitchell. 1995. Masking of children's early understanding of the representational mind: Backwards explanation versus prediction. *Child Development* 66/4: 1022-1039.
- Segal, Gabriel (1998). Representing representations. In Peter Carruthers and Jill Boucher, eds., *Language and thought*, 146-161. Cambridge: Cambridge University Press.
- Slomkowski Cheryl L. and Judy Dunn. 1992. Arguments and relationships within the family: Differences in young children's disputes with mother and sibling. *Developmental Psychology* 28/5: 919-924.
- Tanney, Julia. 2013. *Rules, reason and self-knowledge*. London: Harvard University Press.
- Tesla, Caroline and Judy Dunn. 1992. Getting along or getting your own way: the development of young children's use of argument in conflicts with mother and sibling. *Social Development* 1/2: 107-121.
- Tooming, Uku. 2016. Mental State Attribution for Interactionism. *Studia Philosophica Stonica* 9/1: 184-207.
- Turnbull, Williams, Jeremy Carpendale, and Timothy, O. Racine. 2009. Talk and children's understanding of mind. *Journal of Consciousness Studies* 16/6-8: 140-166.
- Urmson, James O. 1952. Parenthetical verbs. *Mind* 61/244: 480-496.
- Uttich, Kevin and Tania Lombrozo. 2010. Norms inform mental state ascriptions: A rational explanation for the side-effect effect. *Cognition* 116: 87-100.
- Van Cleave, Matthew and Christopher Gauker. 2010. Linguistic practice and false-belief tasks. *Mind and Language* 25/3: 298-328.
- Wellman, Henry R. 2014. *Making minds: How theory of mind develops*. Oxford: Oxford University Press.
- Wierzbicka, Anna. 2006. *English: Meaning and culture*. Oxford: Oxford University Press.
- Wimmer, Heintz and Josef Perner. 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13/1:103-128.
- Wyman, Emily, Hannes Rakoczy, and Michael Tomasello, M. 2009. Normativity and context in young children's pretend play. *Cognitive Development* 24/2: 146-155.
- Zawidzki, Tadeusz W. 2013. *Mindshaping: A new framework for understanding human social cognition*. Cambridge: MIT Press, A Bradford Book.

**VÍCTOR FERNÁNDEZ CASTRO** is a current post-doctoral research fellow at Institut Jean Nicod, (DEC, ENS, PSL Research University). Previously, he obtained his Ph.D. at the University of Granada in 2017, where he was also a post-doctoral research fellow (2017-2018) thanks to the project "Inner speech, Metacognition, and the Narrative View of Identity" (FFI2015-65953-P). His main areas of interest are the philosophy of language and the philosophy of mind and psychology.

**ADDRESS:** Víctor Fernández Castro, Institut Jean Nicod, Département d'Etudes Cognitives, ENS, PSL Research University. 29 rue d'Ulm 75005 Paris, France. vfernandezcastro@gmail.com